

Problem Definition Tools

Dataset

4 Findings

Data Model

Resources

Contents





# Problem Definition

#### **Project Problem**

To predict whether a hotel booking will be cancelled using current and prior booking information.

#### Why is this important?

Throughout the year, hotels get numerous bookings, and many of these bookings are canceled. Cancelled bookings result in a loss of revenue for the company; if management could foresee whether a hotel reservation will be canceled, it would be in their best interests to assign the room to someone else and win their business. This research seeks to use current and historical booking data to predict if a hotel reservation will be canceled.



2

# Tools





#### **Data Processing**



Numpy ,Pandas and,Sklearn

#### Modelling



Scikit-learn

#### visualization



Matplotlib ,Plotly folium and Seaborn





# Dataset

### **Hotel booking Dataset**



	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	meal	country
0	Resort Hotel	0	342	2015	July	27	1	BB	PRT
1	Resort Hotel	0	737	2015	July	27	1	ВВ	PRT
2	Resort Hotel	0	7	2015	July	27	1	ВВ	GBR
3	Resort Hotel	0	13	2015	July	27	1	ВВ	GBR
4	Resort Hotel	0	14	2015	July	27	1	ВВ	GBR
119385	City Hotel	0	23	2017	August	35	30	ВВ	BEL
119386	City Hotel	0	102	2017	August	35	31	ВВ	FRA
119387	City Hotel	0	34	2017	August	35	31	ВВ	DEU
119388	City Hotel	0	109	2017	August	35	31	ВВ	GBR
119389	City Hotel	0	205	2017	August	35	29	НВ	DEU

This data set includes a single file that compares various booking details between two hotels : city hotel and resort hotel.

# Findings

4



## Plot the count plot for both the hotel types





More than 60% of the population booked the City hotel

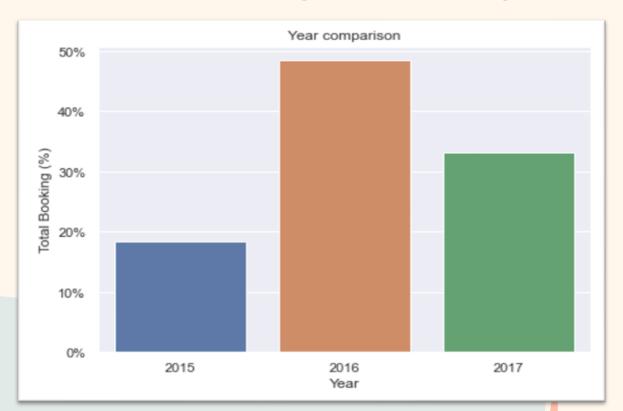


Bookings got canceled 37% of the time. While booking guests did checks-in (did not cancel the booking) almost 63% of the time.

## **How Many Bookings Were Cancelled?**



#### What Is The Percentage Of Booking For Each Year?



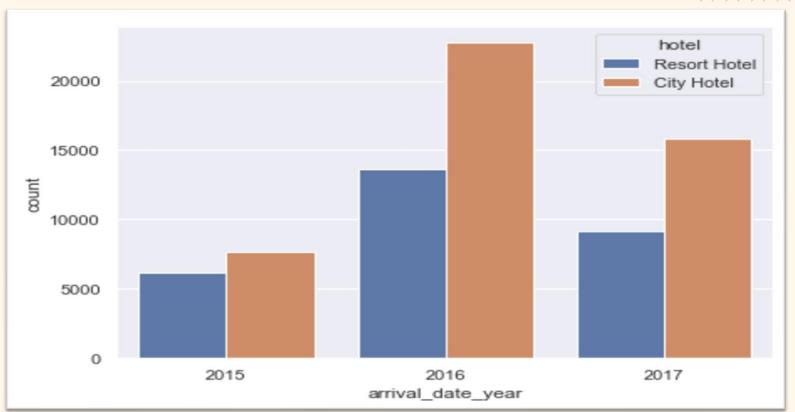
More than double bookings were made in 2016, compared to the previous year.

But the bookings decreased by almost 15% the next year.



## Separate By The Hotel Type

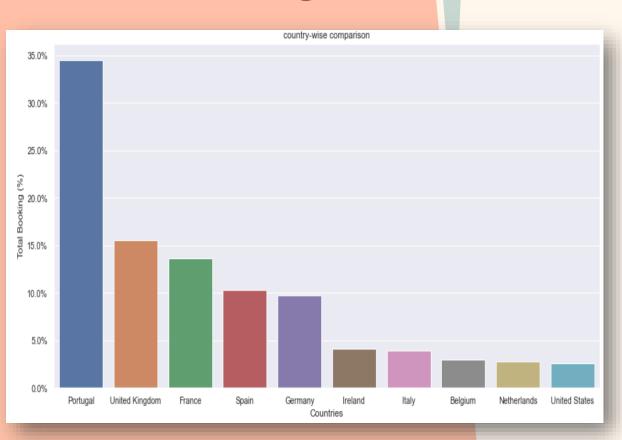




### Which Is The Busiest Month For Hotels?



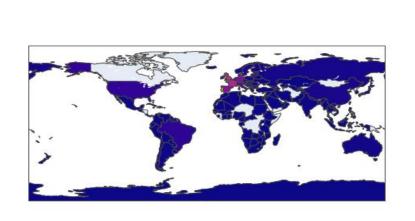
## The most guests are coming??

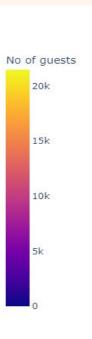


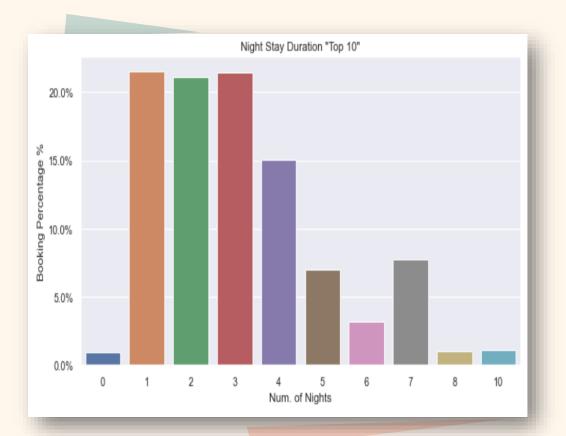
Portugal, UK and France, Spain and Germany are the top countries from most guests come, more than 80% come from these 5 countries.

### The most guests are coming display in map



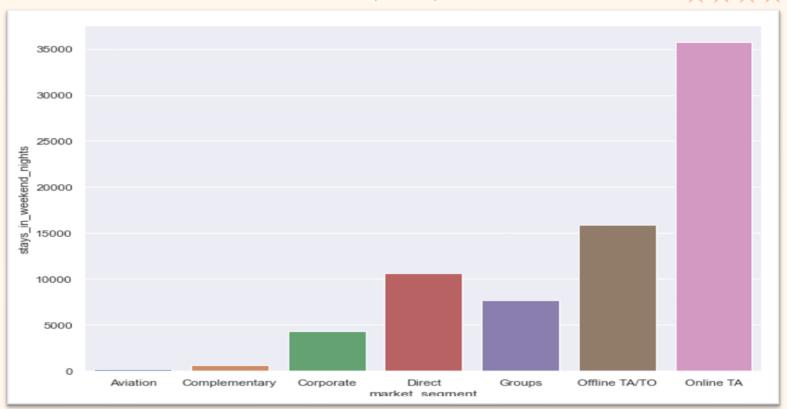




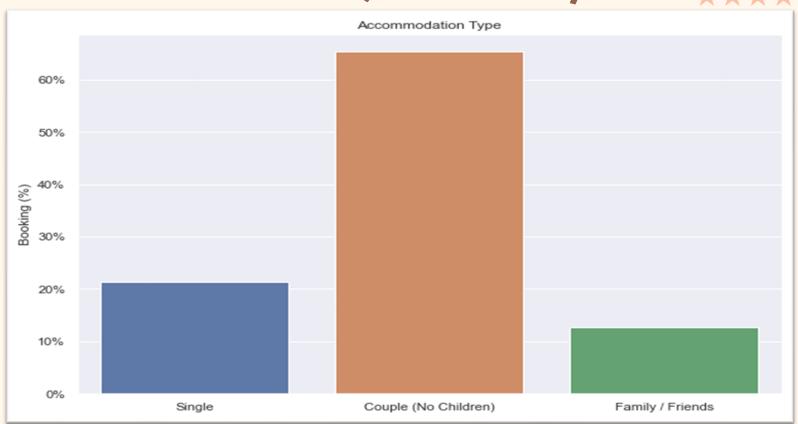


## How Many Days Do People Stay In A Hotel?

# Online agents are the best market for reservation by people



# Customer Type Single, Couple, Multiple Adults, And Family



## Display Heatmap



is_canceled	1	0.29	0.017	0.0081	-0.0061	-0.0018	0.025	0.06	0.005	-0.032	-0.085	0.11	-0.057	-0.14	-0.047	-0.083	0.054	0.048	-0.2	-0.23	- 1.0
lead_time	0.29	1	0.04	0.13	0.0023	0.086	0.17	0.12	-0.038	-0.021	-0.12	0.086	-0.074	0.00015	-0.013	-0.086	0.17	-0.063	-0.12	-0.096	
arrival_date_year	0.017	0.04	1	-0.54	-0.00022	0.021	0.031	0.03	0.055	-0.013	0.01	-0.12	0.029	0.031	0.056	0.034	-0.056	0.2	-0.014	0.11	- 0.8
arrival_date_week_number	0.0081	0.13	-0.54	1	0.067	0.018	0.016	0.026	0.0055	0.01	-0.03	0.036	-0.021	0.0055	-0.018	-0.033	0.023	0.076	0.0019	0.026	
arrival_date_day_of_month	-0.0061	0.0023	-0.00022	0.067	1	-0.016	-0.028	-0.0016	0.015	-0.00023	-0.0061	-0.027	-0.0003	0.011	0.0002	0.0037	0.023	0.03	0.0087	0.0031	
stays_in_weekend_nights	-0.0018	0.086	0.021	0.018	-0.016	1	0.5	0.092	0.046	0.018	-0.087	-0.013	-0.043	0.063	0.16	-0.08	-0.054	0.049	-0.019	0.073	- 0.6
stays_in_week_nights	0.025	0.17	0.031	0.016	-0.028	0.5	1	0.093	0.044	0.02	-0.097	-0.014	-0.049	0.096	0.2	-0.044	-0.002	0.065	-0.025	0.068	
adults	0.06	0.12	0.03	0.026	-0.0016	0.092	0.093	1	0.03	0.018	-0.15	-0.0067	-0.11	-0.052	0.025	-0.17	-0.0083	0.23	0.015	0.12	- 0.4
children	0.005	-0.038	0.055	0.0055	0.015	0.046	0.044	0.03	1	0.024	-0.033	-0.025	-0.021	0.049	0.051	-0.043	-0.033	0.32	0.056	0.082	0.4
babies	-0.032	-0.021	-0.013	0.01	-0.00023	0.018	0.02	0.018	0.024	1	-0.0089	-0.0075	-0.0066	0.083	0.03	-0.0095	-0.011	0.029	0.037	0.098	
is_repeated_guest	-0.085	-0.12	0.01	-0.03	-0.0061	-0.087	-0.097	-0.15	-0.033	-0.0089	1	0.082	0.42	0.012	-0.052	0.16	-0.022	-0.13	0.077	0.013	- o.:
previous_cancellations	0.11	0.086	-0.12	0.036	-0.027	-0.013	-0.014	-0.0067	-0.025	-0.0075	0.082	1	0.15	-0.027	-0.018	-0.0012	0.0059	-0.066	-0.018	-0.048	
previous_bookings_not_canceled	-0.057	-0.074	0.029	-0.021	-0.0003	-0.043	-0.049	-0.11	-0.021	-0.0066	0.42	0.15	1	0.012	-0.046	0.11	-0.0094	-0.072	0.048	0.038	
booking_changes	-0.14	0.00015	0.031	0.0055	0.011	0.063	0.096	-0.052	0.049	0.083	0.012	-0.027	0.012	1	0.036	0.089	-0.012	0.02	0.066	0.053	- 0.
agent	-0.047	-0.013	0.056	-0.018	0.0002	0.16	0.2	0.025	0.051	0.03	-0.052	-0.018	-0.046	0.036	1	-0.12	-0.041	0.017	0.12	0.061	
company	-0.083	-0.086	0.034	-0.033	0.0037	-0.08	-0.044	-0.17	-0.043	-0.0095	0.16	-0.0012	0.11	0.089	-0.12	1	-0.023	-0.13	0.038	-0.091	
days_in_waiting_list	0.054	0.17	-0.056	0.023	0.023	-0.054	-0.002	-0.0083	-0.033	-0.011	-0.022	0.0059	-0.0094	-0.012	-0.041	-0.023	1	-0.041	-0.031	-0.083	
adr	0.048	-0.063	0.2	0.076	0.03	0.049	0.065	0.23	0.32	0.029	-0.13	-0.066	-0.072	0.02	0.017	-0.13	-0.041	1	0.057	0.17	
required_car_parking_spaces	-0.2	-0.12	-0.014	0.0019	0.0087	-0.019	-0.025	0.015	0.056	0.037	0.077	-0.018	0.048	0.066	0.12	0.038	-0.031	0.057	1	0.083	
total_of_special_requests	-0.23	-0.096	0.11	0.026	0.0031	0.073	0.068	0.12	0.082	0.098	0.013	-0.048	0.038	0.053	0.061	-0.091	-0.083	0.17	0.083	1	
	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	is_repeated_guest	previous_cancellations	revious_bookings_not_canceled	booking_changes	agent	company	days_in_waiting_list	adr	required_car_parking_spaces	total_of_special_requests	-

05

# Data Model



#### **□** Logistic Regression Model

	Not Canceled = 0	Canceled =1
precision	0.98	0.96
Recall	0.99	0.94
F1-score	0.98	0.95



#### **☐** KNN Model

	Not Canceled = 0	Canceled =1
precision	0.78	0.48
Recall	0.87	0.33
F1-score	0.82	0.39

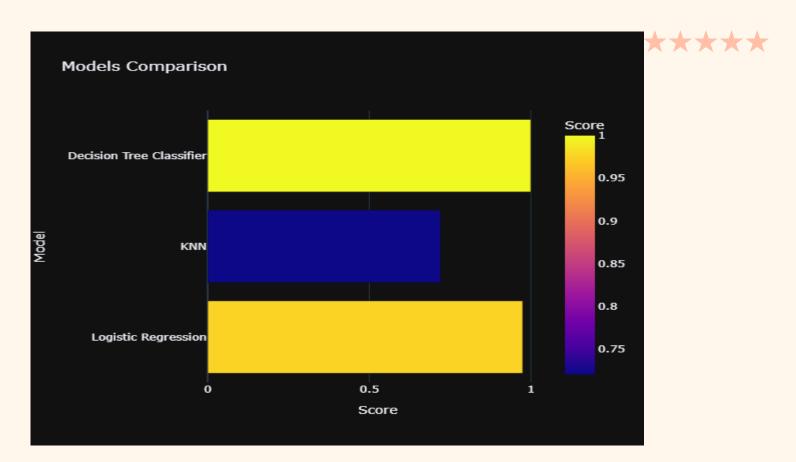
#### **□** Decision Tree Model

	Not Canceled = 0	Canceled =1
precision	0.98	0.96
Recall	0.99	0.94
F1-score	0.98	0.95



#### **Comparison Accuracy**

	Model	Score
1	Decision Tree Classifier	1.00
2	Logistic Regression	0.97
3	KNN	0.72







- 1. <a href="https://www.kaggle.com/jessemostipak/hotel-booking-demand">https://www.kaggle.com/jessemostipak/hotel-booking-demand</a>
- 2. Exploratory Data Analysis of the Hotel Booking Demand with Python | by Aaqib Qadeer Soomro | Analytics Vidhya | Medium

# Thanks ...

Do you have any questions?